© 2018 IJSRSET | Volume 4 | Issue 4 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

An Efficient Approach for Accurate Frequent Pattern Mining Practising Threshold Values

Janak Thakkar*[1], Dr. Mehul Parikh[2]
*[1]IT Department, R C Technical Institute, Ahmedabad, Gujarat, India
[2]Associate Professor, IT Department, LDCE, Ahmedabad, Gujarat, India

ABSTRACT

Frequent pattern mining is the highest priority research area for researchers in data mining. It covers major real world applications. Many algorithms, techniques and tools are available for frequent pattern mining with popular Apriori and FP growth. Many algorithms have been designed to mine frequent closed item sets using single minimum support value approaches and various data structures. However, it is necessary to improve the existing methods in terms of execution time and memory consumption as well as the reduced item sets. Along with this our paper represent the new method to generate frequent closed item set. The proposed method is based on the concept of data set reduction but with preserving the interesting and important patterns. Proposed method is focused with the concept of dual minimum threshold values. The experimental results have shown that the proposed methodology is not only generating the closed item sets with reduced data set but also less mining time and memory space.

Keywords: Data Mining, Frequent Pattern Mining, Closed Item Sets, Close Frequent Patterns, Minimum Support Threshold

I. INTRODUCTION

In today's world, there is rapid growth in producing data in the world. It is not possible to analyze the data with the old and conventional method. Due to this limitation, a large amount of data is still kept without used. Data mining is a technique which comes into the picture to solve the problem by mining the important and useful information from the large dataset. The rapid growth in database and the usage of that data for corporate sector, government offices, and scientific organizations and marketing world.

Frequent patterns are the important itemsets from the large datasets, which occurs with greater frequency than the user-specified threshold. Frequent patterns plays key role in mining associations, correlations, and many other interesting relationships among data in datasets. It can also be utilised in data indexing, classification, clustering, and other data mining tasks as well. Closed frequent sets are important in the time line series, future predictions and frequency based approximation. Thus, FPM is an important data mining technique and high priority topic for the research.

II. RELATED WORK

Many algorithms have been proposed for the Frequent Item set mining. Agrawal etal.[1,2] were pioneer for the FPM with concept of the market basket analysis, which focus on the association rule mining, by which interesting relationships from the large data sets can be found out. For example - bread and butter. In the year 2016 Kamlesh Aahuja and Sarika Jain [3] proposed the method for the mining closed itemsets with single minimum support threshold. many algorithms like MS-apriori, MSB-

apriori , MSB-apriori+[4] and OSPF - Growth+[5] algorithms on the basis of the multiple minimum support value. UFAE algorithm uncertain filtering and extending proposed by the s k jain and goyal [6] for mining top k rank patterns from the uncertain large datasets. Close and Closet+[9] were proposed to produced the closed item sets from the large datasets. Many more algorithm like DITP Miner[8] for mining inter transactional patterns, dNC_ECPM[10] for mining erasable unwanted closed patterns. Focus of the algorithms to reduce the data set to generate the frequent patterns in different ways, but sometimes it discards the important profitable patterns.

III. PROBLEM DEFINATION

After literature review it was noticed that by ding scan with the single minimum support or by single minimum support threshold we can reduce the data set and generate the frequent patterns but sometimes we lose the important FP which can be more profitable and accurate for the future predictions.

Old approach was creating the closed itemset but on the bases of threshold values so it loses some important FPs. Also it results in creating more mining time as well as more memory usage because it creates two different trees to sort out the selected FPs.

IV. PROPOSED SOLUTION

The proposed methodology applies two different threshold values in which one is user defined and another is calculated on the bases of the transaction frequency.

Input : Transactional Data Set, Multiple Support Threshold

Step1: Scan the transaction database D to find out the number of occurrences of all size 1 itemsets.

Step 2: The support of each item is computed & stored in a data structure. This data structure has two parts: the head part & the body part. The body of this data structure contains all single items with their support. we count each item's support by using compressed data structure.

Step 3: All the size 1 items are arranged in the decreasing order of their support count. This is the candidate set of all size 1 items

Step 4: the support of each size 1 item is compared with the minimum threshold value

Step 5: Eliminate all those size-1 itemsets of step 1 whose support is less than the MST. It will result in a compressed table, which consists of all the frequent items of size 1

Step 6: Now sort all the itemsets of last step in descending order of their item count (frequency)

Step 7: Create a 2 dimensional data structure and store the transaction and the correspondent frequency in that table

Step 8: Then scan the data structure created in step 7 to locate all the K size itemsets. calc mst, Select only those item sets whose support is greater than the minimum threshold (MST).

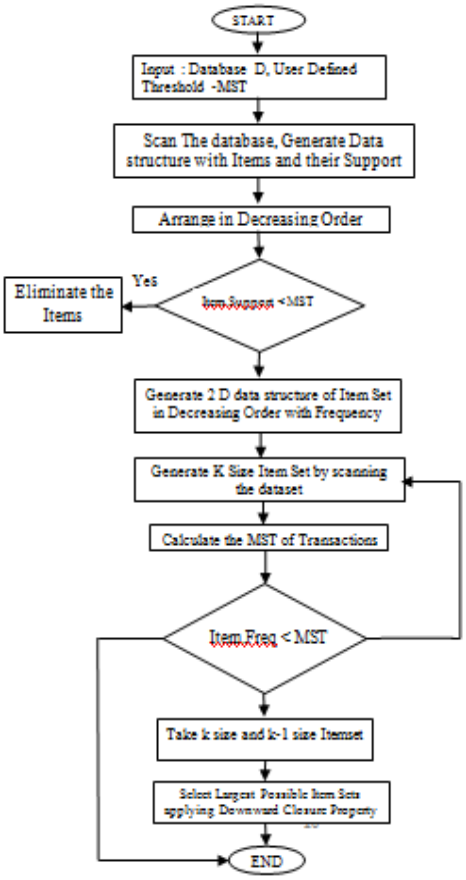
Step 9: If the support count of the k size item sets is less than the MST then take K size itemsets and k – 1 size itemsets together to generate a k - 1 size item set. Continue this step until no item set having support greater than the MST is found.

Step 10: All the largest possible size item sets are found in step 7 then by applying the downward closure property, All the subsets found are also frequent.

Step 11: If no such transaction exists in table of step 7 then go to step 12 otherwise repeat step 7 to 10.

Step 12: exit.

V. FLOW OF THE PROPOSED WORK



VI. EXPERIMENTAL RESULTS

Results are tested on dataset of super market. This data set of super market consisting of the large data of the transaction of the customers. First proposed algorithm generate the Size - 1 item set with the user defined threshold value, which is 40% in our case. In the process of step 6 to step 11 frequent closed item set were generated with the minimum threshold support value calculated on the bases of the transactional frequency of the transactions. To generate the closed frequent itemsets the k size and k-1 size item set are together and k-1 size item set generated.

Figure 1, figure 2 and figure 3 shows the comparison of the mining time required for mining the dataset, comparison of the frequent closed item sets generated and the comparison of the memory usage respectively. The results are the runtime comparison between the old methodology and the proposed methodology.

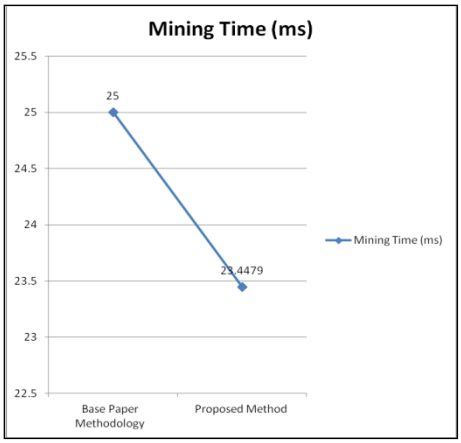


Figure 1. Comparison of Mining Time

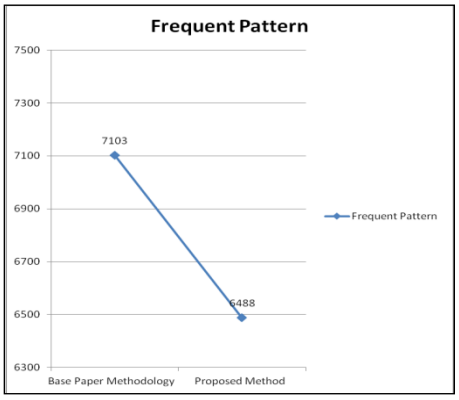


Figure 2. Comparison of Frequent Pattern Generated

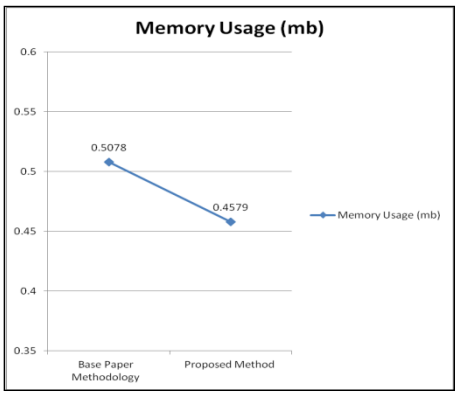


Figure 3. Comparison of Memory Usage

VII. CONCLUSION AND FUTURE WORK

The purpose of the research is to generate the closed item set which are not only frequent but more accurate for utilization. Also we were intended to find out such an algorithm which can be memory efficient and also with less mining time. The proposed methodology serves all the objectives in

most efficient way with less mining time and memory without losing the important patterns.

VIII. REFERENCES

[1]. First Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACMSIGMOD international conference on management of data(SIGMOD'93), pages207–216.1993

[2]. Agrawal R, Srikant R. Fast algorithms for mining association rules. InProceedings of the 1994 international conference on very large databases(VLDB'94), pages487–499.1994

[3]. "Data Elimination Based Technique for Mining Frequent Closed Item Set" by Kamlesh Ahuja and Sarika Jain, Year 2017, IEEE

[4]. Mining Frequent Patterns with Multiple Minimum Supports using Basic Apriori" by Tiantian Xu, Xiangjun Dong, Year 2013 in Ninth International Conference on Natural Computation (ICNC)

[5]. "An optimized frequent pattern mining algorithm with multiple minimum supports" by Hsiao-Wei Hu, Hao-Chen Chang, Wen-Shiu Lin, Year 2016 in IEEE International Conference on Big Data (Big Data)

[6]. An Efficient Algorithm For Mining Top-Rank-K Frequent Patterns From Uncertain Databases" by Neha Goyal and S K Jain,Year 2016, IEEE

[7]. An Efficient Algorithm For Mining Top-Rank-K Frequent Patterns From Uncertain Databases" by Neha Goyal and S K Jain,Year 2016, IEEE

[8]. "An Improved Algorithm for Mining Frequent Inter - Transaction Patterns" by Thanh-Ngo Nguyen, Loan T.T. Nguyen, Ngoc- Thanh Nguyen, Year 2017 European Union Journals

[9]. "A Review paper for mining Frequent Closed Itemsets" by Dungarval Jayesh and Neeru Yadav, Year 2014, International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 1, January 2014 pg. 534-537

[10]. "Efficient algorithms for mining erasable closed patterns from product datasets" by Tuong Le, Giang Nguyen, Tzung-Pei Hong, Year 2016 in IEEE Access Journal

[11]. "An Improved and Efficient Frequent Pattern Mining Approach to Discover Frequent Patterns among Important Attributes in Large Data set Using IA-TJFGTT" by Saravanan. Suba, Dr.T. Christopher, Year 2016 in IEEE International Conference on Advances in Computer Applications (ICACA)

[12]. "A Survey Paper For Finding Frequent Pattern In Text Mining", Ms. Sonam Tripathi , Asst Prof. Tripti Sharma, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 3, March 2015

[13]. "Frequent Itemset Mining in Data Mining: A Survey", Rana Ishita, Amit Rathod, International Journal of Computer Applications (0975-8887) Volume 139

[14]. "Frequent Itemset Mining Algorithms :A Literature Survey", O.Jamsheela, Raju.G,Year 2015 in IEEE Journal

[15]. " Frequent Pattern Mining under Multiple Support Thresholds" by SADEQ DARRAB, BELGIN ERGENÇ in WSEAS TRANSACTIONS on COMPUTER RESEARCH, 2016

[16]. "An efficient algorithm for mining maximal frequent patterns over data streams" by Junrui Yang , Yanjun Wei , Fenfen Zhou in 7th International Conference on Intelligent Human-Machine Systems and Cybernetics, 2015

[17]. "Frequent Pattern Generation in AssociationRule Mining using Weighted Support"by Subrata Bose and Subrata Datta, year 2015, IEEE